How many mailings are enough?

Vehovar, Vasja; Lozar, Katja

How Many Mailings Are Enough?

Vasja Vehovar and Katja Lozar

Abstract: *This paper addresses the balance between costs and errors in mail surveys. Two alternatives are compared: a larger sample with less follow-up mailings and a smaller sample with more follow-up mailings. The comparison involves a detailed elaboration of mean squared errors and cost functions. Based on the model, the key variables are discussed. The empirical example refers to the percentage of companies with access to the Internet. It is shown that the above elaboration can be helpful for practical decisions.*

Keywords: *survey costs, nonresponse, mail surveys*

1 Introduction

In survey research, we often discuss various procedures for improving the quality of data but rather rarely do we discuss survey costs. However, when quality improvement efforts are discussed in such an isolated form a heavy mismatch between theory and practice may occur (Groves 1989, vi, vii).

In this paper, we pose the following practical question: What is the optimum balance between the errors and the costs of a mail survey?

We use the general understanding of survey errors and survey costs (Kish 1965, Groves 1989). Specifically, we narrow our analysis down to the issues of sample size and number of contacts. Both parameters, large initial sample size and large number of follow-ups, lead to smaller error, but at the same time, they also produce higher costs. We search for the precise balance between survey costs and errors.

Empirically, we present the case of the Total Design Method (TDM) mail survey (Dillman 1978) with three follow-ups (the third conducted by telephone). The key question is whether or not to use the third follow-up. The dilemma can be expressed in the following question: Is it better to have a small initial sample with three follow-ups, or a large initial sample with only two follow-ups? We explore the following factors: the expected nonresponse conversion rate after the 3rd follow-up, the relative costs for the 3rd follow-up and the relative bias after the 2nd follow-up. We also outline some general guidelines for finding the optimal number of contacts.

2 Errors

The components of survey errors have already been well-elaborated (Groves 1989, 3). However, in this paper we limit ourselves to the sampling error and to the component that belongs to the nonresponse bias. We use the standard form of the mean square error (Kish 1965):

$$MSE\,(p_c) = E\{p_c - P\}^2 = E\{p_c - E(p_c)\}^2 + \{E(p_c) - P\}^2,$$

where P is a true population value and p_c is a sample estimate where c runs across all possible samples. The first part represents the variance part and the second part represents the bias part.

The sampling error can be regulated with an increase/decrease of the sample size. The (nonresponse) bias component of the MSE can be, at least in this context, reduced with additional contacts. We will assume a simple random sample (SRS). We will further assume that we are dealing with a simple population parameter - the population percentage P. In our specific example the root mean square error (RMSE) is based on the sum of variance and squared bias:

$$RMSE(p) = \sqrt{MSE(p)} = \sqrt{Var(p) + Bias(p)^2}\ .$$

We have an estimate of *RMSE(p)* for a population percentage *P*:

$$rmse(p) = \sqrt{\frac{p(1-p)}{n} + (P-p)^2}\ ,$$

where we omitted the finite population correction. Of course, in the case of continuous variable y we have:

$$rmse\left(\bar{y}\right) = \sqrt{\frac{s^2(y)}{n} + \left(\bar{Y} - \bar{y}\right)^2}\ .$$

The above definition of RMSE(p) is not taking into account the different number of follow-ups. In the case of v follow-ups we have[1] the following expression for *rmse(p)*:

$$rmse(p)=\sqrt{\frac{p(1-p)}{n}+(P-p)^2}=\sqrt{\frac{1}{\sum_{i=0}^{v} n_i'}\left[\frac{\sum_{i=0}^{v} p_i n_i'}{\sum_{i=0}^{v} n_1'}*\left(1-\frac{\sum_{i=0}^{v} p_i n_i'}{\sum_{i=0}^{v} n_1'}\right)\right]+\left[P-\frac{\sum_{i=0}^{v} p_i n_i'}{\sum_{i=0}^{v} n_i'}\right]^2}\;.$$

Here we have n'_i as the achieved sample size in the i-th follow-up, p_i as the estimate at the i-th follow-up, where i=1... v. We have n_i' as the function of the initial sample size n^* and the completion rate at the i-th follow-up, $CR_i = n_i'/n^*$. The RMSE function can be thus rewritten:

$$rmse(p)=\sqrt{\frac{1}{n\sum_{i=0}^{v} CR_i}\left[\frac{\sum_{i=0}^{v} p_i * CR_i}{\sum_{i=0}^{v} CR_i}*\left(1-\frac{\sum_{i=0}^{v} p_i * CR_i}{\sum_{i=0}^{v} CR_i}\right)\right]+\left[P-\frac{\sum_{i=0}^{v} p_i * CR_i}{\sum_{i=0}^{v} CR_i}\right]^2}\;.$$

RMSE is thus a function of the population percentage P, nonresponse bias, completion rates, initial sample size n^* and number of follow-ups v. We have - at least in this context – no influence on the value of the population percentage P nor can we regulate the sample estimate p (i.e. the bias). We also have no influence on the completion rate at the i-th follow-up (*CRi*). On the other hand, we can regulate the initial sample size n^* and the number of follow-ups v. In the expressions above we have assumed that all the previous nonrespondents have been included in each additional follow-up.

3 Costs

As we have already mentioned, the changes in the sample size and the number of follow-ups influence the survey costs. The cost function in our example is thus not a continuous one as in the case of one initial contact. In a simplified form it can be written:

$$T = K + C_v = K + (K_v + C_{vn}) = K + (v+1)*A + n^* \sum_{i=0}^{v+1} (1 - CC_i) * c_i \;,$$

[1] Since v is the number of follow-ups the first contact has no follow-up, therefore v = 0.

where the total costs (T) of the survey consist of constant costs (K) and variable costs (C_v). The constant costs include: design of the survey, construction of a questionnaire, data management, data analysis, preparation, printing as well as circulation of survey reports, overhead costs etc. The variable costs have two components:

- the costs K_v that vary with the number of follow-ups (but not with the initial sample size). We assumed that these costs are the same for each follow-up, therefore we can express K_v as the product of number of phases in the survey research process and a certain level of administrative costs $((v+1)*A)$.
- the variable costs C_{vn} that are proportional to the initial sample size n^* and to the variable costs at the i-th follow-up. These costs are the function of the cumulative contact rate (CCi) and ci - the costs per unit within each follow-up[2]. These costs include the costs of paper, envelopes, printing, and a part of the administrative work that depends on the number of mailings (packing etc.). Of course, $CC_0 = CC_1=0$, since all units receive the initial mailing and the reminder. The last component $(i=v+1)$ does not refer to any follow-up but includes only the costs of data entry. Therefore, CCi takes value $CC_{v+1} =(1- n/n^*)$, so that $n^*(1- CC_{v+1})$ gives the number of all responding units which are denoted as n.

4 Optimisation

The aim of the optimal design may be stated in two alternative ways: achieving minimum MSE for fixed costs, or achieving minimum costs for fixed MSE. Both principles would generally lead to the same solution (Kish 1965, 263-264). However, unlike with the standard sampling theory, it is difficult to find the analytical solution when the variable to optimise is a discrete one – the number of contacts v.

In the case of the above two equations (costs, RMSE) we have calculate costs and RMSE for each value of v and then compare the values.

We will concentrate on the optimisation of the RMSE for fixed costs. It is possible to increase the sampling error (with a decrease of n^*, initial sample size) and simultaneously reduce the nonresponse error (with an increase of v, number of contacts), or the other way around, but the total costs must remain the same. Typically, we can have a large initial sample and a small number of follow-ups, or the opposite, a small initial sample and a large number of follow-ups.

[2] At this we have in mind the initial sample since survey costs comprise costs for respondents and nonrespondents.

5 Example

In 1996, a mail survey on the use of Internet among Slovenian companies was conducted as a part of a larger research project *Research on Internet in Slovenia* at the Faculty of Social Sciences (http://www.ris.org). The design was a standard TDM design. The envelopes containing a cover letter, a questionnaire and a return envelope were sent to 4,698 companies. A week later a reminder was sent. After two weeks, the second follow-up with a replacement questionnaire and a return envelope was sent to the nonrespondents. Three weeks later CATI (computer assisted telephone interviewing) follow-up was performed among the nonrespondents. The response rate after three follow-ups was 67.9% and the completion rate was 60.0%. In reality, only 10% of the nonrespondents were contacted by telephone. However, in the calculations below we assumed that all the nonrespondents were surveyed. Without this simplification the model would become much more complicated as we would introduce another variable in order to optimise - the sub-sampling rate for the third follow-up.

The key population parameter in the survey was a percentage of companies with the access to Internet. Since the true population value was unknown we have assumed that the true value was the value achieved after the last follow-up. The bias after the 3rd follow-up was therefore automatically set to zero.

Table 1: Optimum design in a mail survey

No. of follow-ups	estimate p of P	costs for a unit of init. sample	cumul. costs for a unit of initial s.	% of total costs	cumul. % of total costs	Cumulat-ive compl. rate	initial sample size	final sample size (resp.)	rmse(p) - estimate of RMSE
0	18.0%	47.5	47.5	23.6	23.6	18.2%	19,880	3,618	0.040506
1	16.6%	26	73.5	12.9	36.5	26.5%	12,848	3,405	0.026770
2	14.1%	35.5	109	17.7	54.2	44.2%	8663	3,829	0.005712
3	14.0%	92	201	45.8	100.0	60.0%	4698	2,819	0.006535

We can observe that after three follow-ups 14.0% of companies had access to Internet. Without the telephone follow-up the estimate would be 14.1%. Of course, these are cumulative percentages; the percentage for the 3rd wave respondents alone is lower than 14.0%. If we performed only one follow-up the estimated percentage would be 16.6%, and without any follow-ups the estimate would be 18.0%. Obviously, a strong nonresponse bias exists, however the decision concerning the optimal number of contacts is not that obvious at all, unless we perform some calculations.

All designs in Table 1 assume the same fixed budged. For the budget needed for three follow-ups and an initial sample of 4,698 units, we could omit the telephone follow-up and enlarge the initial sample size to 8,663. If we omitted the 3rd and the 2nd follow-up, we could have an initial sample of 12,848 companies. In case of no follow-up we could send the mailing to 19,880 companies. In all these cases the bias and the sample variance vary considerable. However, the optimum for this fixed budget is achieved in the case of two follow-ups (the smallest RMSE).

6 Generalisation

The example above was, no doubt, a very specific one. However, a general solution can not be derived analytically, so we would like to find the basic principles by varying the parameters in the above example. For this purpose we will alter one of the variables while keeping the others constant. We will concentrate on the decision whether to use the 3rd follow-up or not. We will thus compare the situation after the second and after third follow-up. Three key parameters are important in this process:

1. relative costs of the third follow-up in comparison to the costs of the first two follow-ups,
2. bias after the 2nd follow-up,
3. the nonresponse conversion rate in the 3rd follow-up.

A) The costs for the 3rd follow-up

How high can the costs of the 3rd follow-up be (in relation to the costs of previous contacts), so that the use of 3rd follow-up would reach the optimal RMSE?

If we have the same fixed costs for both situations, with two and with three follow-ups, this will obviously create a difference in the initial sample size. Therefore, the important variable here is the sample size, which is fixed at 4,698 for two follow-ups. That is why the corresponding RMSE line is **constant** for two follow-ups. Of course with a 3rd follow-up the RMSE changes according to the relative costs of the 3rd follow-up. From Figure 1 we can see that RMSE is smaller for 3 then for 2 follow-ups only if the costs for the 3rd follow-up are less then 40% of the costs for two follow-ups. The 3rd follow-up is thus optimal only if its costs will not increase the previous costs for more then 40%[3].

[3] In our case two follow-ups was the best decision, because costs for the 3rd follow-up presented more then 84% of the costs for first two follow-ups.

Figure 1: Influence of the relative costs of the 3^{rd} follow-up[4]

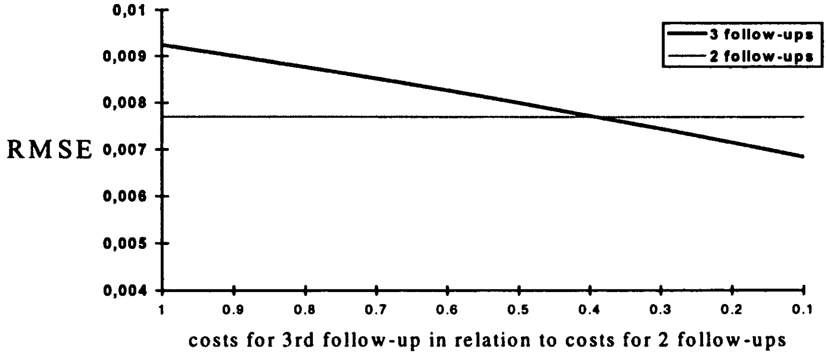


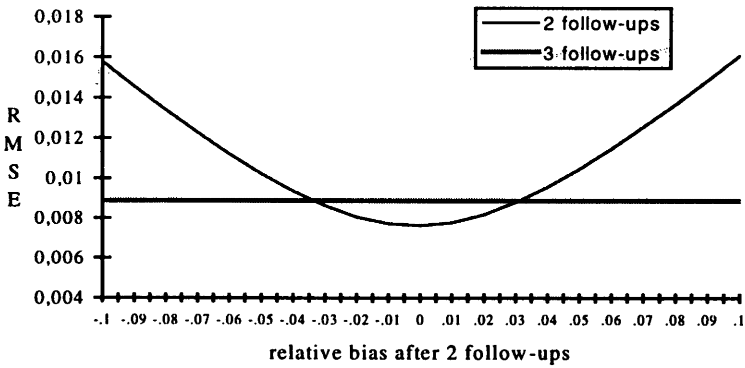
Figure 1 presents only the relationship when a relative bias after the 2^{nd} follow-up takes a specific value of 1%. If the relative bias after the 2^{nd} follow-up is larger, the line presenting the corresponding RMSE will be higher. In such a case the 3^{rd} follow-up can be reasonable also in the case of higher relative costs.

B) The bias after two follow-ups

How large should the relative bias be after two follow-ups in order to justify the use of an additional contact? In this situation the RMSE for the 3^{rd} follow-up is **constant**, as we assume no bias after the last follow-up.

4 The constant factors in this case are: costs for two follow-ups - 109 SIT, estimated percentage after 2^{nd} follow-up - 0.141, estimated percentage after 3^{rd} follow-up - 0.140, relative bias after 2^{nd} follow-up - 0.7%, completion rate after 2^{nd} follow-up - 0.442, completion rate after 3^{rd} follow-up - 0.600.

Figure 2: Influence of the relative bias after two follow-ups[5]



The variable factor is the relative bias after two follow-ups. It changes from -10% to +10%. We can observe that RMSE is smaller for three then for two follow-ups if the relative bias after two follow-ups is larger than 0.035[6].

Again, the above figure shows only a specific case where the costs for the 3rd follow-up represent 84% of the costs for previous follow-ups. If the costs for the 3rd follow-up would be smaller, the initial sample size for 3 follow-ups could be larger - the horizontal line would be lower, with different intercepts.

C) The nonresponse conversion rate after the 3rd follow-up

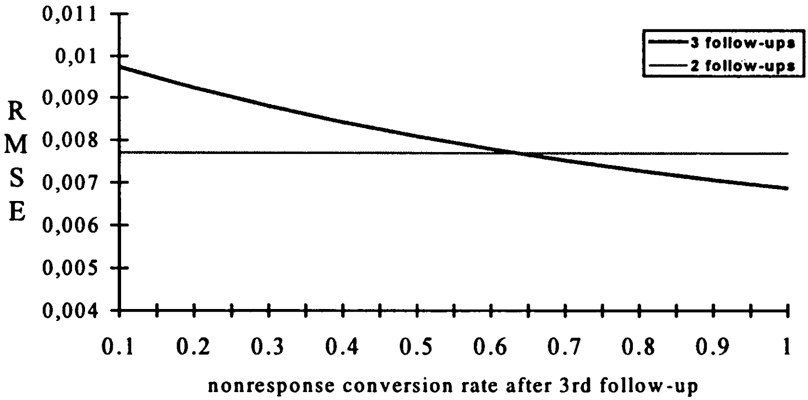
Another factor that can influence the decision as regards the 3rd follow-up is the nonresponse conversion rate after the 3rd follow-up. How large should this conversion be in order to use the 3rd follow-up?

---

5 The constant factors in this case are costs for two follow-ups - 109 SIT, costs for three follow-ups: 210 SIT, initial sample size for two follow-ups – 4698, initial sample size for three follow-ups - 2548. In this case the total costs for 2 or 3 follow-ups are the same. The completion rate after the 2nd follow-up - 0.442, completion rate after the 3rd follow-up - 0.600, estimated percentage after 3rd follow-up - 0.141, relative bias after 3rd follow-up - 0.

6 In our case the relative bias after two follow-ups was less then 1%, so the 3rd follow-up was not needed.

Figure 3: The impact of the expected nonresponse conversion rate[7]



The variable factor is the nonresponse conversion rate after the 3[rd] follow-up. This rate influences the total completion rate and therefore also the final sample sizes. The larger the conversion rate, the larger the final sample and smaller sampling variance and RMSE. Of course, this variable factor has no impact on the completion rate in case of two follow-ups, so this line is a **constant**.

We can see that RMSE is smaller for three then for two follow-ups when nonresponse conversion rate after the 3[rd] follow-up is larger then 60%[8].

7 Conclusion

We have demonstrated the impact of various factors effecting the decision of whether or not to use a third follow-up contact in a mail survey. The same principles apply to any previous or additional follow-up. Of course, the relationship between the parameters involved is complex and depends on many specific circumstances. Other factors also may

7 The constant factors in this case are costs for two follow-ups - 109 SIT, costs for three follow-ups - 210 SIT, initial sample size for two follow-ups – 4698, initial sample size for three follow-ups – 2548, estimated percentage after 2 follow-ups - 0.141, estimated percentage after 3 follow-ups - 0.140, completion rate after 2 follow-ups - 0.442. In this case total costs for 2 or 3 follow-ups are the same.

8 In our case the nonresponse conversion rate after the 3[rd] follow-up was 24.8% what increased the completion rate from 44.2% to 60%. As the nonresponse conversion rate was small, the 3[rd] follow-up was not worthwhile to use.

play an important role, such as time constraints or low quality of late responses[9]. However, when faced with a clear dilemma between sampling error and nonresponse bias the above results can be useful.

For the simultaneous understanding of all three factors together (bias, costs, response) a multivariate presentation in a three-dimensional space may be helpful. In such a space a sort of pyramid can be drawn. Only within the body of such a pyramid can the parameters take on values that may justify the use of the third follow-up.

It is somewhat difficult to perform the above calculation in practice. One obvious complication is the case of different sub-populations which behave differently. Another obstacle may be that we have no information about the bias and the nonresponse rates. There may even be difficulties with the accurate anticipation of the costs. Of course, in such situations a good decision cannot be reached. However, it is reasonable to make certain estimates from previous surveys or, at least an educated guess. It is also possible to make estimates from earlier stages of the same survey. Based on these assumptions we can - with the aid of the above-described model - obtain a better understanding of the interaction between costs and errors in mail surveys.

References

Dillman, D.A. (1972). Increasing Mail Questionnaire Response in Large Samples of the General Public. Public Opinion Quarterly, 36, 2, pp. 254-257

Dillman D.A. (1978). Mail and Telephone Survey. The Total Design Method. New York: John Wiley and Sons

Dillman, D.A. (1989). Our Changing Sample Survey Technologies. Choices, 4, 4, pp. 40-41

Dillman, D.A. (1991). The Design and Administration of Mail Surveys. Annual Review of Sociology, 17, pp. 225-249

Dillman, D.A., Clark, J.R. and Treat, J.B. (1994). Influence of 13 Design Factors on Completion Rates to Decennial Cencus Questionnaires. Paper presented at the 1994 Annual Research Conference of the U.S. Bureau of the Cencus, Arlington, Virginia, March 21, 1994

Green, K.E. (1991). Reluctant Respondents: Differences Between Early, Late, and Nonresponders to a Mail Survey. Journal of Experimental Education, 59, 3, pp. 268-276

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: John Wiley and Sons

[9] Several authors recognise that late respondents are less interested in the survey topic and less willing to cooperate in the survey (Green 1991, 268-276; Kojetin et al. 1993, 838-843).

Kish, L. (1965). Survey Sampling. New York: John Wiley and Sons

Kojetin, B.A., Borgida, E. and Snyder, M. (1993). Survey Topic Involvement and Nonresponse Bias. Proceeding of the Section on Survey Research Method, Volume II., American Statistical Association, pp. 838-843

Martin, C.L. (1994). The Impact of Topic Interest on Mail Survey Response Behaviour. Journal of the Market Research Society, 36, 4, pp. 327-338

Pearl, D.K. and Fairley, D. (1985). Testing for the Potential for Nonresponse Bias in Sample Surveys. Public Opinion Quarterly, 49, 4, pp. 553-560

Roeher, G.A. (1963). Effective Techniques in Increasing Response to Mailed Questionnaires. Public Opinion Quarterly, 27, 2, pp. 299-302